Finite Temperature String Method with Umbrella Sampling: Application on a Side Chain Flipping in Mhp1 Transporter

Hyun Deok Song and Fangqiang Zhu*

Department of Physics, Indiana University - Purdue University Indianapolis, IN, USA

ABSTRACT

Protein conformational change is of central importance in molecular biology. Here we demonstrate a computational approach to characterize the transition between two metastable conformations in all-atom simulations. Our approach is based on the finite temperature string method, and the implementation is essentially a generalization of umbrella sampling simulations with Hamiltonian replica exchange. We represent the transition pathway by a curve in the conformational space, with the curve parameter taken as the reaction coordinate. Our approach can efficiently refine a transition pathway and compute a one-dimensional free energy as a function of the reaction coordinate. A diffusion model can then be used to calculate the forward and backward transition rates, the major kinetic quantities for the transition. We applied the

* Email: fzhu0@iupui.edu; Phone: (317) 274-7634

This is the author's manuscript of the article published in final edited form as:
Song, H. D., & Zhu, F. (2017). Finite Temperature String Method with Umbrella Sampling: Application on a Side Chain Flipping in Mhp1 Transporter. The Journal of Physical Chemistry B, 121(15), 3376–3386. https://doi.org/10.1021/acs.jpcb.6b08568

approach on a local transition in the ligand-free Mhp1 transporter, between its outward-facing conformation and an intermediate conformation with the side chain of Phe305 flipped to the outside of the protein. Our simulations predict that the flipped-out position of this side chain has a free energy 6.5 kcal/mol higher than the original position in the crystal structure, and that the forward and backward transition rates are in the millisecond and sub-microsecond time scales, respectively.

INTRODUCTION

Conformational changes in biomolecules such as proteins are of central importance in molecular biology. For many proteins such as membrane transporters, conformational changes are indispensable steps in their functioning cycles. In the typical two-state model, a protein has two alternative metastable conformations and may make spontaneous transitions between the two conformational states in equilibrium. The forward and backward transition rates are the major quantitative properties of the transition, and the ratio of the two rates also determines the equilibrium probabilities of the two conformations. Because conformational transitions are rare random events and occur very fast in comparison to the much longer times spent in each metastable state, they are normally difficult to directly detect in experiments. All-atom molecular dynamics (MD) simulations are a powerful tool to reveal detailed mechanism in protein systems. As a visionary leader in this field, Klaus Schulten pioneered the innovative applications of MD simulations to tackle many interesting biophysical problems.[1] Naturally, the MD technique also has the potential capability to elucidate protein conformational transitions.

Ideally, sufficiently long MD simulations, in which a large number of transitions between the two conformational states occur spontaneously, can reproduce the equilibrium ensemble of the protein and reveal all thermodynamic and kinetic properties of the conformational transition. Unfortunately, although this straightforward approach is conceptually simple and robust, it requires extremely long simulation times in practice. Even with the most powerful computational resource nowadays, the currently affordable simulation times are only sufficient for systematically characterizing small proteins with relatively fast transition rates.[2, 3] Nevertheless,

the equilibrium ensemble is an important concept, and all alternative techniques discussed below are designed to reproduce properties of this unbiased ensemble.

To overcome the problem of insufficient simulation times, a large variety of methods, often collectively called enhanced sampling techniques, have been developed. These methods include transition path sampling[4], metadynamics[5], accelerated MD[6], adaptive biasing force[7], milestoning[8], dynamic importance sampling[9], weighted ensemble[10], steered MD[11], among many others. All of the enhanced sampling techniques introduce some form of bias in the simulations, aiming to calculate properties such as the equilibrium probabilities and the transition rates from the biased simulations of affordable times. In particular, the string method[12, 13] has been demonstrated to be applicable to large systems with many relevant degrees of freedom. One variant of the technique, the finite temperature string method[14, 15], can be used to obtain a one-dimensional (1D) free energy profile as a function of the reaction coordinate,[16] thus providing the equilibrium probabilities of each metastable conformation.

The original implementation[14] of the finite temperature string method employed constrained simulations in the hyperplanes to sample the conformational space, similar to the "locally updated planes" technique[17]. Alternatively, umbrella sampling with harmonic potentials on the reaction coordinate were also applied to calculate the free energy.[18] Since then, Voronoi tessellation with reflective boundaries[15, 16] has been established as a common implementation for the finite temperature string method. In a recent study of an adenylate kinase[19], we adopted an implementation which essentially amounts to umbrella sampling (with Hamiltonian replica exchange[20]) along a curve in the multidimensional conformational space, and obtained a free energy profile along a predetermined transition pathway.[19] In this study, we further present

approaches to refine the conformational transition pathway by such simulations, and apply the techniques to elucidate the transition of a side chain flipping in the Mhp1 transporter.

Mhp1 is a bacterial secondary transporter that cotransports Na^+ ions and hydantoin molecules.[21] High-resolution crystal structures[22-24] are available for two major conformational states of Mhp1: the outward-facing (OF) and the inward-facing (IF) conformations. Several MD studies[23-28] were reported for the dynamics and conformational changes of Mhp1. In our recent attempt to characterize the transition between the OF and the IF conformations of Mhp1, we identified two inter-helical loops that exhibit slow transitions in their secondary structures.[29] Since then, we have identified another slow degree of freedom involving local structures: the flipping of the side chain of Phe305 in TM8. As shown in Fig. 1, the phenylalanine side chain is in the interior cavity on the extracellular side of the protein in the OF structure[22] but points outward in the IF structure[23]. Our previous equilibrium simulations[29] indicated that the positions of Phe305 in the respective crystal structures are quite stable, without undergoing any spontaneous flipping. In a preliminary simulation when we drove the protein backbone (by restraints) from the OF to the IF state, the Phe305 side chain did not flip either, and thus always remained in the interior cavity. In the end of this simulation, consequently, even though the protein backbone was similar to the target IF crystal structure, the Phe305 side chain did not reach the correct position. Furthermore, this side chain appears to have an effect on the stability of the protein, especially for the IF conformation with the extracellular cavity largely closed, because the cavity with the Phe305 inside would be too crowded. According to these findings, the flipping of the Phe305 side chain is a necessary step in the complete conformational transition of Mhp1, but the spontaneous flipping would be very rare within typical MD times.

When all heavy atoms of Mhp1 were subject to the restraints in our previous targeted molecular dynamics (TMD) simulations[29], the Phe305 side chain was indeed enforced to flip. Particularly, in the TMD simulation[29] that drove the protein from the OF to the IF conformation, the Phe305 side chain was flipped to the outside before all other major changes in the protein conformation started. Reversely, in the IF-to-OF TMD simulation[29], the flipping of the side chain into the interior cavity occurred after all other major changes had completed. Therefore, both TMD simulations[29] consistently indicated that the flipping of the Phe305 side chain is the first step in the OF-to-IF transition. In this study, we apply our method of umbrella sampling with Hamiltonian replica exchange[20] to characterize the transition between the OF state and a metastable state denoted as OF$^{'}$ in which the side chain of Phe305 is located outside of the interior cavity (see Fig. 1).

THEORY AND METHODS

In this section, we first formulate the general theory and techniques to calculate the conformational free energy and the transition rates. We then describe our MD simulations for the transitions between the OF and the OF' conformations of the ligand-free Mhp1 protein.

Coarse coordinates and configuration space

In general, protein conformations are described by some chosen "coarse coordinates", which can be either Cartesian coordinates or collective variables[13]. Suppose N such coarse coordinates are chosen, collectively denoted by an N-dimensional vector $\vec{X} \equiv (X_1, X_2, ..., X_N)$. Each unique set of the coarse coordinates thus corresponds to one point in an N-dimensional configuration

space. Although N can be large, it is typically much smaller than the dimensionality of the phase space that represents the exact system microstates (i.e., the positions and momenta of all atoms in the system). Using $p(X_1,...,X_N)$ to denote the multivariate probability distribution for $\vec{X}$ in the equilibrium ensemble at a constant temperature T, a corresponding multidimensional free energy, denoted by $G(X_1,...,X_N)$ or $G(\vec{X})$, can be defined as a function of these coarse coordinates:

$$G(X_1,...,X_N) \equiv -k_B T \cdot \ln p(X_1,...,X_N) + const., \qquad (1)$$

where k_B is the Boltzmann constant. In principle, the equilibrium probability $p(X_1,...,X_N)$ and the free energy $G(X_1,...,X_N)$ can be directly obtained from sufficiently long unbiased simulations, although in practice the affordable simulation times are normally far from sufficient. The conventional string method[13] was designed to calculate the values of $G(X_1,...,X_N)$ on a pathway that connects two regions in the configuration space, each corresponding to a metastable conformation.

We note that in almost all enhanced sampling techniques such as the string methods, the bias is applied on the chosen coarse coordinates only. Other degrees of freedom not represented by the coarse coordinates will not be facilitated by the enhanced sampling, and consequently a proper sampling of those degrees of freedom relies completely on spontaneous relaxation. If some of the unrepresented degrees of freedom involve infrequent transitions between multiple states, they may suffer poor sampling within the simulation time, thus resulting in hysteresis.[29, 30] Therefore, it would be preferable to select a relatively large number of coarse coordinates, as in

many applications of the string methods[12-16], to reduce the risk of ignoring slow and relevant degrees of freedom.

One-dimensional free energy as a function of curve parameter

As mentioned earlier, the conventional string method[12, 13] aims to calculate the values of $G(X_1,...,X_N)$ on a minimum free energy pathway, which is a curve in the multidimensional configuration space formed by the coarse coordinates. This free energy profile, however, does not reflect values of $G(X_1,...,X_N)$ in any off-pathway region of the configuration space. In contrast, the finite temperature string method[14, 15] calculates a 1D free energy that incorporates the probabilities both on and off the pathway, as described below.

In general, a continuous curve in the configuration space can be mathematically described as $\vec{X}^{\,\mathrm{cur}}(\alpha)$, in which any given value of the curve parameter α corresponds to a unique point on the curve. We use $s(\alpha)$ to denote the cumulative curve length, such that the arc length between two points $\vec{X}^{\,\mathrm{cur}}(\alpha_1)$ and $\vec{X}^{\,\mathrm{cur}}(\alpha_2)$ on the curve is given by

$$s(\alpha_2)-s(\alpha_1)=\int_{\alpha_1}^{\alpha_2}\left|d\vec{X}^{\,\mathrm{cur}}(\alpha)\right|=\int_{\alpha_1}^{\alpha_2}\left|\frac{d\vec{X}^{\,\mathrm{cur}}(\alpha)}{d\alpha}\right|d\alpha\,. \qquad (2)$$

Without loss of generality, we further assume that the curve is parameterized such that the magnitude of the derivative above is a constant L at any α:

$$\left|\frac{d\vec{X}^{\,\mathrm{cur}}(\alpha)}{d\alpha}\right|=L\,. \qquad (3)$$

Such uniform parameterization ensures that the arc length is a linear function of the curve parameter: $s(\alpha_2) - s(\alpha_1) = L \cdot (\alpha_2 - \alpha_1)$.

We may project any point $\vec{X}$ in the configuration space onto the given curve $\vec{X}^{\,\mathrm{cur}}(\alpha)$ by applying an operator P_α:

$$P_\alpha(\vec{X}) \equiv \arg\min_{\alpha} \left| \vec{X}^{\,\mathrm{cur}}(\alpha) - \vec{X} \right|. \qquad (4)$$

The operator takes $\vec{X}$ as the input and returns the value of the curve parameter α that corresponds to the point on the curve with the shortest distance to $\vec{X}$. We assume that the curvature of the curve is sufficiently low, such that the projection of any point in the accessible space (i.e., regions with nontrivial probabilities in the configuration space) onto the curve is unique, and that an infinitesimal change of $\vec{X}$ will not result in a finite jump in $P_\alpha(\vec{X})$. By such projection, any protein conformation can be mapped to a single value of α, and the curve parameter α can be used as the reaction coordinate for the conformational transition. We may thus take α as a single random variable and define its probability distribution $p(\alpha)$ in the equilibrium ensemble, which corresponds to a 1D free energy:

$$G(\alpha) \equiv -k_B T \cdot \ln p(\alpha) + const. \qquad (5)$$

Whereas the conventional string method[12, 13] calculates the multidimensional free energy $G(X_1,...,X_N)$ on the pathway curve alone, the 1D free energy $G(\alpha)$ here effectively incorporates degrees of freedom in the perpendicular dimensions as well. The value of $G(\alpha)$ at a given α does not describe the probability for merely the point $\vec{X}^{\,\mathrm{cur}}(\alpha)$ on the curve, but is actually determined by the integrated probability over the hypersurface in the configuration

space that projects to that point. The 1D free energy profile $G(\alpha)$ thus faithfully describes the thermodynamics of the entire accessible conformational space mapped to the reaction coordinate. For the typical scenario of a two-state transition, the profile would feature two valleys (local minima) separated by a barrier. The equilibrium probability of each metastable conformation can also be directly obtained from $G(\alpha)$.

Umbrella sampling along a curve

As mentioned earlier, the finite temperature string method could be implemented in simulations in which the coarse coordinates $\vec{X}$ are constrained at hyperplanes[14] or confined in Voronoi cells with reflective boundaries[15, 16]. Alternatively, similar to some previous studies,[18] we recently adopted an implementation[19] based on the umbrella sampling with Hamiltonian replica exchange[20] to calculate $G(\alpha)$. This approach is described below.

As in the typical umbrella sampling, our method employs a set of M simulations, or M umbrella windows. The simulation in each umbrella window is subject to a biasing potential that restrains the reaction coordinate α to the vicinity of a reference value. Typical umbrella potentials on the reaction coordinate are of the harmonic form:

$$U_i(\alpha) = \frac{K_\alpha}{2}\left(\alpha - \alpha_i^{\text{ref}}\right)^2, \qquad (6)$$

in which K_α is the spring constant, and α_i^{ref} is the reference value and the center of the harmonic potential $U_i(\alpha)$ for window i, with i =1,…, M. The restraint on the coarse coordinates is therefore

$$U_i(\vec{X}) = \frac{K_\alpha}{2}\left[P_\alpha(\vec{X}) - \alpha_i^{\text{ref}}\right]^2. \qquad (7)$$

The harmonic biasing potentials above are very common in the umbrella sampling. In our case, however, the exact evaluation of $P_\alpha(\vec{X})$ and the gradients requires nonlinear optimization and can be computationally expensive. We therefore invoke a local linear approximation to simplify the implementation of the umbrella potential. Because each window only samples a narrow range of α due to the restraint, we do a linear expansion of the curve near the reference value α_i^{ref}:

$$\vec{X}^{\text{cur}}(\alpha_i^{\text{ref}} + \Delta\alpha) \approx \vec{X}^{\text{cur}}(\alpha_i^{\text{ref}}) + \Delta\alpha \cdot \frac{d}{d\alpha}\vec{X}^{\text{cur}}(\alpha_i^{\text{ref}}). \qquad (8)$$

We further denote

$$\vec{X}_i^{\text{ref}} \equiv \vec{X}^{\text{cur}}(\alpha_i^{\text{ref}}) \qquad (9)$$

as the reference point on the curve for window i, and define

$$\hat{r}_i \equiv \frac{d}{d\alpha}\vec{X}^{\text{cur}}(\alpha_i^{\text{ref}}) \Big/ L \qquad (10)$$

as the curve direction at $\vec{X}_i^{\text{ref}}$, in which L according to Eq. 3 is the magnitude of the derivative and is a constant given the uniform parameterization of the curve. Therefore, $\hat{r}_i$ is a unit vector:

$$\left|\hat{r}_i\right| = 1. \qquad (11)$$

With the definitions above, Eq. 8 may be written as

$$\vec{X}^{\text{cur}}(\alpha_i^{\text{ref}} + \Delta\alpha) \approx \vec{X}_i^{\text{ref}} + (L\Delta\alpha) \cdot \hat{r}_i. \qquad (12)$$

Eq. 12 represents a straight line passing through $\vec{X}_i^{\,\mathrm{ref}}$ along the tangent direction $\hat{r}_i$. Under this approximation, the projection $\alpha_\perp = P_\alpha(\vec{X})$ becomes a linear operation, and the point on the line with the shortest distance to $\vec{X}$ is:

$$\vec{X}^{\,\mathrm{cur}}(\alpha_\perp) = \vec{X}^{\,\mathrm{ref}} + \left[(\vec{X} - \vec{X}_i^{\,\mathrm{ref}}) \cdot \hat{r}_i\right]\hat{r}_i . \qquad (13)$$

In comparison to Eq. 12, the projection operator can then be evaluated as

$$P_\alpha(\vec{X}) = \alpha^{\mathrm{ref}} + (\vec{X} - \vec{X}^{\,\mathrm{ref}}) \cdot \hat{r}_i / L . \qquad (14)$$

The umbrella potential in Eq. 7 thus becomes

$$U_i(\vec{X}) = \frac{K}{2}\left[(\vec{X} - \vec{X}_i^{\,\mathrm{ref}}) \cdot \hat{r}_i\right]^2, \qquad (15)$$

in which

$$K \equiv K_\alpha / L^2 \qquad (16)$$

can be taken as the spring constant on the coarse coordinates. From the trajectories of simulations under these restraints, the free energy $G(\alpha)$ can be calculated using the weighted histogram analysis method (WHAM).[30, 31] One could also implement Hamiltonian replica exchange[20] in these simulations to enhance the sampling efficiency.

Unlike the restraints in the conventional string method which depend on the distance between $\vec{X}$ and $\vec{X}_i^{\,\mathrm{ref}}$, our restraint here (Eq. 15) only depends on the component of $\vec{X}$ along the tangent direction of the curve. The components perpendicular to the tangent are not subject to any bias. The restraint essentially forces the coarse coordinates $\vec{X}$ to be near a hyperplane (of

$N-1$ dimensions) that orthogonally intersects the pathway curve at $\vec{X}_i^{\text{ref}}$, but allows free sampling within the hyperplane. Consequently, the 1D free energy $G(\alpha)$ calculated from such simulations will have effectively integrated out (marginalized) all degrees of freedom perpendicular to the curve direction.

Iterative refinement of transition pathway

The umbrella sampling above requires a predetermined pathway curve. In practice, however, a proper transition pathway is not known a priori, and has to be gradually refined from some rough initial guess. Specifically, starting from an initial pathway, we may take M uniformly spaced points $\vec{X}_i^{\text{ref}}$ on the curve along with the tangent vectors at each point, and carry out one round of umbrella sampling simulations as described earlier. From the simulation trajectories, we will calculate the average coordinates $\vec{X}_i^{\text{mean}}$ (i =1,…, M) for each of the M windows. In comparison to the $\{\vec{X}_i^{\text{ref}}\}$ on the original curve, the set of new coordinates $\{\vec{X}_i^{\text{mean}}\}$ is expected to be energetically more favorable. We will thus fit a smooth curve through $\{\vec{X}_i^{\text{mean}}\}$ by multidimensional curve fitting techniques[19] derived from the harmonic Fourier beads method[32], and use the fitted curve as an improved pathway. The new pathway curve will then be further sampled and refined in the next round of simulations. The refinement can be repeated until the pathway becomes stabilized over the iterations.

The procedure of pathway refinement here is similar to a previous application[33] of the conventional string method. There is a major difference, however, in that the restraints[33] in the conventional string method are applied on all components of the coarse coordinates and will thus restrict the deviation of $\{\vec{X}_i^{\text{mean}}\}$ from $\{\vec{X}_i^{\text{ref}}\}$. In contrast, the 1D restraints here do not impose

restrictions on the directions perpendicular to the curve, thus allowing full relaxations along those directions. Consequently, the transition pathway is expected to evolve faster over the iterations toward the converged curve.

As discussed earlier, we take the curve parameter for the transition pathway as the reaction coordinate for the transition. We note that in principle, the ideal reaction coordinate should be directly related to the committor.[34] Here we do not seek to identify the perfect reaction coordinate, and we will consider it satisfactory as along as the calculations based on the pathway curve could reasonably predict the thermodynamics and kinetics of the transition. We also note that the minimum free energy pathway in the conventional string method[13] has the requirement that in the Cartesian space of the underlying atomic coordinates, the tangent direction of the curve should be parallel to the mean force.[13] We do not impose this requirement for the transition pathway here. In fact, the pathway curve in our method merely serves as a projection reference such that the accessible configuration space can be mapped to a single reaction coordinate; the curve itself does not need to closely represent the trajectories of typical spontaneous transitions in the configuration space.

Calculation of transition rates based on a diffusion model

In a typical two-state model, metastable conformations A and B correspond to two major minima separated by a barrier in the 1D free energy profile $G(\alpha)$. Suppose that the two minima are at reaction coordinates α_A and α_B, respectively, and that the peak of the free energy barrier is at α^*, with $\alpha_A < \alpha^* < \alpha_B$. At equilibrium, the probabilities of finding the protein conformation in each state, P_A and P_B (with $P_A + P_B = 1$), can be predicted from the free energy:

$$P_A = \frac{\int_{-\infty}^{\alpha^*} \exp[-G(\alpha)/k_B T]d\alpha}{\int_{-\infty}^{\infty} \exp[-G(\alpha)/k_B T]d\alpha}, \qquad (17a)$$

$$P_B = \frac{\int_{\alpha^*}^{\infty} \exp[-G(\alpha)/k_B T]d\alpha}{\int_{-\infty}^{\infty} \exp[-G(\alpha)/k_B T]d\alpha}. \qquad (17b)$$

Although the free energy describes the thermodynamics of the protein conformations, it is not sufficient to determine the kinetic rates of the transition. Because the protein dynamics can be reduced to the time evolution of a single reaction coordinate α, we may describe the conformational transitions as a 1D diffusion along the reaction coordinate, quantified by the diffusion coefficients $D(\alpha)$. In general, $D(\alpha)$ could depend on the value of α, analogous to the position-dependent diffusion coefficients for real particles. The diffusion coefficients can be estimated from the umbrella sampling simulations under harmonic biasing potentials. Specifically, from a single simulation with α restrained in a narrow range, the local diffusion coefficient can be calculated as[35]

$$D = \mathrm{var}(\alpha)/\tau, \qquad (18)$$

in which var(α) denotes the variance of α, and τ is the autocorrelation time constant for the equilibrium trajectory $\alpha(t)$ under the harmonic potential:

$$\tau \equiv \int_0^{\infty} \langle \delta\alpha(0) \cdot \delta\alpha(t) \rangle dt \Big/ \mathrm{var}(\alpha), \qquad (19)$$

with $\delta\alpha(t) \equiv \alpha(t) - \langle \alpha \rangle$.

Suppose we have uniformly recorded n data points with a time interval Δt from a simulation trajectory $\alpha(t)$ with a total length $T = n \cdot \Delta t$, and denote these data as $\alpha_i \equiv \alpha(i\Delta t)$ (i =1,…, n). If the autocorrelation function (as in the integrant of Eq. 19) vanishes at a time scale much shorter than T, the time constant τ can be evaluated as[36]

$$\tau \approx \left[\frac{n \cdot \mathrm{var}(\overline{\alpha})}{\mathrm{var}(\alpha)} - 1 \right] \Delta t / 2 , \tag{20}$$

where $\overline{\alpha} \equiv \sum \alpha_i / n$ is the average of the data from the trajectory, and its variance var($\overline{\alpha}$) can be estimated by block averages.[37] If n is large, the equation can be further approximated by

$$\tau \approx \frac{T}{2} \cdot \frac{\mathrm{var}(\overline{\alpha})}{\mathrm{var}(\alpha)} . \tag{21}$$

By combining the values of D calculated for each umbrella window, we can then obtain the position-dependent diffusion coefficients $D(\alpha)$.

With the obtained free energy $G(\alpha)$ and diffusion coefficients $D(\alpha)$, the dynamics of the reaction coordinate can be described by a Smoluchowski equation:

$$\frac{\partial}{\partial t} p(\alpha, t) = -\frac{\partial}{\partial \alpha} j(\alpha, t) , \tag{22a}$$

$$j(\alpha, t) = -D(\alpha) e^{-G(\alpha)/k_B T} \frac{\partial}{\partial \alpha} [e^{G(\alpha)/k_B T} p(\alpha, t)] , \tag{22b}$$

in which $p(\alpha, t)$ and $j(\alpha, t)$ are the time-dependent probability density distribution and the flux along the reaction coordinate, respectively.

At equilibrium, the net flux is clearly zero everywhere, and the equilibrium probability distribution is given by Eq. 5. To calculate the transition rates, however, here we only focus on those systems in the transition region which have most recently visited state A rather than state B. For such systems in the equilibrium ensemble, there exists a stationary flux from A to B despite the time-independent probability densities. In fact, this constant flux is equal to the rate (k_0) of spontaneous transitions from A to B in the equilibrium ensemble. The rate can thus be calculated from the stationary solution of Eq. 22 under proper boundary conditions, resulting in

$$k_0 = \frac{1}{\left\{\int_{-\infty}^{\infty} d\alpha \exp[-G(\alpha)/k_B T]\right\} \left\{\int_{\alpha_A}^{\alpha_B} d\alpha \exp[G(\alpha)/k_B T]/D(\alpha)\right\}}. \qquad (23)$$

Similarly, the rate of spontaneous transitions from state B to A is also equal to k_0 in the equilibrium ensemble. Scaled by the equilibrium probabilities P_A and P_B (Eq. 17), the forward and backward transition rates are then

$$k_{A\to B} = \frac{k_0}{P_A} = \frac{1}{\left\{\int_{-\infty}^{\alpha^*} d\alpha \exp[-G(\alpha)/k_B T]\right\} \left\{\int_{\alpha_A}^{\alpha_B} d\alpha \exp[G(\alpha)/k_B T]/D(\alpha)\right\}}, \qquad (24a)$$

$$k_{B\to A} = \frac{k_0}{P_B} = \frac{1}{\left\{\int_{\alpha^*}^{\infty} d\alpha \exp[-G(\alpha)/k_B T]\right\} \left\{\int_{\alpha_A}^{\alpha_B} d\alpha \exp[G(\alpha)/k_B T]/D(\alpha)\right\}}. \qquad (24b)$$

Therefore, the major kinetic constants for the transition, $k_{A\to B}$ and $k_{B\to A}$, can be calculated from the diffusion model with the free energy and the diffusion coefficients obtained from the simulations.

Details of Mhp1 simulations

Simulation system. The protein in our simulation consists of residue 13 to 470 in Mhp1, with the initial OF structure taken from PDB 2JLN[22]. We adopted the standard protonation states at pH 7 for all residues, with all His residues protonated at the ε position. The protein was embedded in a bilayer of 184 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) lipid molecules, and solvated by 13,060 water molecules. Five Cl^- ions were added to render the system electrically neutral. The entire system consists of 69,342 atoms in total. This setup is similar to the one in our earlier study.[29]

Simulation protocols. All simulations were performed under the periodic boundary conditions and using the CHARMM (Ver. c36) force field for the protein (with the CMAP correction)[38-40] and lipids[41], the TIP3P water model[42], and the NAMD2 (Ver. 2.11) program[43], with a time step of 2 fs. All bond lengths involving hydrogen atoms were constrained using the SHAKE[44] and SETTLE[45] algorithms. We adopted a cutoff distance of 12 Å for nonbonded interactions, with a smooth switching function taking effect at 10 Å. Full electrostatics was calculated every 4 fs using the particle-mesh Ewald method[46]. A constant temperature of 300 K was maintained by Langevin dynamics with a damping coefficient of 0.1 ps^{-1}. A constant pressure of 1 atm was achieved using the Nose-Hoover Langevin piston method[47], with the size of the periodic box allowed to fluctuate but the lengths in *x* and *y* kept equal to each other. In the equilibrated systems, the periodic box has dimensions of ~82 Å × ~82 Å × ~100 Å.

Coarse coordinates. To describe the transition of the Phe305 side chain, we chose the C_4 atom (i.e., the distal carbon atom in the phenyl ring) of that residue as a representative atom. Because Met178 has major interactions with Phe305, we also included the C_γ atom (which is bonded to the C_β and the sulfur atoms) in the Met178 side chain as a representative atom. In addition, we

took the backbone C_α atoms of some nearby residues, namely, residues 168-184 in TM5 and residues 296-325 in TM8, as representative atoms as well. In total, we thus have 49 atoms, or 147 Cartesian coordinates, to represent the local conformation around Phe305. The coarse coordinates $\vec{X} \equiv (X_1, X_2, ..., X_N)$ are then defined from these Cartesian coordinates as $X_i = w_i x_i$, in which x_i is a Cartesian coordinate and w_i is the corresponding weight. In this study, we assigned a weight of 10 $Å^{-1}$ for the coordinates of the C_4 atom in Phe305, and a weight of 1 $Å^{-1}$ for all other coordinates. The coarse coordinates here are thus dimensionless. The OF and OF' states have a distance of ~80 in the configuration space formed by these coarse coordinates.

Because our coarse coordinates are based on Cartesian rather than internal coordinates, we applied strong harmonic restraints[19, 29] on the center (with a spring constant of 1,000 kcal/mol/$Å^2$) and the orientation angle (200 kcal/mol/degree^2) of the entire protein (represented by the C_α atoms of residues 20 to 466) in the umbrella sampling simulations, to eliminate the overall rigid-body translation and rotation. With the protein center and orientation fixed, the coarse coordinates here thus represent the local structures around Phe305 in relation to the entire protein.

Initial coordinates. To generate initial coordinates for umbrella sampling, we performed a simulation of 8 ns starting from the OF structure. In this simulation, the dihedral angle formed by the N-C_α-C_β-C_4 atoms in Phe305 was gradually driven counterclockwise from 167° as in the OF structure[22] to 266° (or -94°) as in the IF structure[23], by a moving harmonic restraint of 200 kcal/mol/rad^2. The restraint enforced a flipping of the Phe305 side chain, thus with the OF' state obtained. The snapshots in the simulation trajectory were then used as the initial coordinates in the subsequent umbrella sampling simulations.

Phe305 is in proximity to the EL4 loop that precedes TM8, and changes in the secondary structure of the loop are important steps to further reach the IF conformation.[29] Because this study is focused on the transitions between the OF and the OF' states only, it is not desirable to have spontaneous transitions (albeit very rare) from OF' to other intermediate states toward IF. We therefore applied additional restraints to prevent major changes in the EL4 loop structure in all simulations here. Specifically, flat-bottomed harmonic potentials with spring constant 0.1 kcal/mol/degree^2 were applied to keep the backbone φ angles of Gly292 and Gly293 within the range from 10° to 180° and the φ angle of Val294 within the range from -180° to -10°, in consistence with their values in the OF state[22]. These restraining potentials are zero if the torsions are in the specified ranges and only act on the torsions when they deviate from those ranges.

Pathway refinement. As described earlier, the refinement of the transition pathway is performed in iterations. In each iteration, a pathway curve is fitted[19] through the average coarse coordinates from the simulations in the previous iteration. Then a set of umbrella sampling simulations along the pathway is carried out, thus generating new average coarse coordinates for fitting the pathway in the next iteration. In this study, we employ M = 64 windows in the umbrella sampling. The applied umbrella potential in each window is in the form of Eq. 15, with the spring constant K = 0.3 kcal/mol for the coarse coordinates. The first and last umbrella windows, however, are treated differently and are subject to half-harmonic potentials instead, thus allowing unbiased sampling beyond the two ends of the curve. Hamiltonian replica exchange[20] was implemented to facilitate convergence in the umbrella sampling, allowing the swap of two simulations that sample adjacent windows, as similarly done in our previous studies[19]. We

performed eight iterations of the pathway refinement, each of which was run for 5 ns per umbrella window, followed by one more iteration with 10 ns per window.

Production run. After the iterative refinement and equilibration above, we performed a last round of longer simulations to calculate the free energy along the final pathway. The simulation setup, with umbrella sampling and Hamiltonian replica exchange[20], was identical to that in the pathway refinement described above, except that each umbrella window was simulated for 100 ns here. Trajectories of the last 80 ns were used in WHAM[26, 27] to calculate the free energy. Because the calculation of diffusion coefficients based on Eq. 18 is only valid for continuous trajectories, we further extended the umbrella sampling simulations by another 5 ns without the Hamiltonian replica exchange.

Unbiased simulations at the barrier top. After the peak (α^*) of the free energy profile was identified, we carried out unbiased simulations starting at α^* to evaluate the quality of the reaction coordinate and to estimate an effective local diffusion coefficient. Specifically, we first took 10 frames from the trajectories of the umbrella windows that sampled the vicinity of α^*. For each frame, we restrained the reaction coordinate at α^* in a simulation of 1 ns, using a strong harmonic potential in the form of Eq. 15 with K = 10 kcal/mol. From its trajectory, we then took 10 snapshots for the coordinates and velocities of all atoms in the system. A total of 100 such snapshots were thus obtained from the 10 restrained simulations above. Next, we generated a "conjugate" for each snapshot by copying the coordinates and inverting (i.e., changing the sign of) the velocities for all the atoms. We thus have a total of 200 snapshots, or 100 pairs of conjugate snapshots with identical coordinates but opposite velocities. Using these snapshots as the initial coordinates and velocities, we carried out 200 independent simulations, each of which lasted for 1 ns and was completely unbiased without any restraints.

RESULTS

As described in Methods, we implemented the finite temperature string method in the form of umbrella sampling with Hamiltonian replica exchange[20] to characterize the flipping transition of the Phe305 side chain in Mhp1. The transition pathway for these simulations connects the OF and the OF' states (see Fig. 1). In the OF state, the Phe305 side chain is in the interior cavity and surrounded by other hydrophobic residues such as Leu175, Met178 and Phe267, whereas in the OF' state the side chain is flipped to the outside and makes direct contacts with the hydrophobic tails of some lipid molecules. The curve parameter α of the pathway was taken as the reaction coordinate for the transition. The trajectories from the umbrella sampling simulations recorded the intermediate conformations along the transition pathway.

A visual inspection of the simulation trajectories reveals that some conformations on the same hyperplane (i.e., with common values of the reaction coordinate) are actually not similar to each other, and instead belong to highly distinct modes. Figure 2 displays the position of the C_4 atom of Phe305 relative to the side chain of Met178 in all frames of the simulation trajectories. To make transitions between the OF and the OF' states, the Phe305 side chain needs to go from one side of Met178 to the other. On some hyperplanes, there appears to be two different conformational modes, with the side chain of Phe305 "below" or "above" Met178 (see Fig. 2), respectively. Conformations in the former mode are connected to both the OF and the OF' structures, and thus are potentially real transition states. Conformations in the latter mode, in contrast, are quite similar to the OF' structure except for some tilting of the TM8 helix toward TM5. Importantly, these conformations are connected to the OF' state only but cannot directly

make a further transition to the OF state. Clearly, conformations in this mode are not intermediates in spontaneous transitions between the OF and the OF' states, and we thus refer to them as "false transition states" here. Unfortunately, the reaction coordinate cannot distinguish the true and false transition states in this case. However, in the analysis we could make corrections to the free energy by discounting the false transition states. Specifically, in the simulation trajectories, among all the frames on a given hyperplane with reaction coordinate α, we counted the number of frames in the false transition states. The fraction of such frames was then taken as an estimate for the equilibrium conditional probabilities $p(\mathit{false}\,|\,\alpha)$ for having a false transition state at the given α. Obviously, this probability may vary with the hyperplane (described by α).

The 1D free energy $G_0(\alpha)$ calculated by WHAM[26, 27] is shown in Fig. 3A (*dashed* line). As demonstrated in Ref. 30, $G_0(\alpha)$ can also be very well approximated by a coarse-grained free energy obtained by integrating the mean restraining forces. Therefore, the uncertainty in the mean forces can be used to estimate the statistical errors in $G_0(\alpha)$.[30] Using this method[30], from the variances of the mean restraining force in each umbrella window, the cumulated statistical error in $G_0(\alpha)$ is estimated to be 0.11 kcal/mol. Furthermore, we calculated the inconsistency coefficients θ (defined in Ref. 30) for every pairs of histograms in adjacent umbrella windows. A θ value much larger than 1 would indicate a major inconsistency between the two histograms.[30] As shown in Fig. 3C, the θ values for our histograms are all below 1, thus reporting no anomaly. Therefore, our results for $G_0(\alpha)$ appear to be consistent and accurate.

The calculated $G_0(\alpha)$ above is directly related to the equilibrium probability distribution of the reaction coordinate α. However, as discussed earlier, a portion of the probabilities in the

barrier region results from false transition states which cannot directly evolve to the OF state due to the wrong side chain mode. To discount these false transition states, we made a correction for the free energy in the barrier region:

$$G(\alpha) = G_0(\alpha) - k_B T \ln[1 - p(\,false \mid \alpha)]\,. \qquad (25)$$

The barrier in the corrected free energy $G(\alpha)$ (Fig. 3A, *solid* line) is thus related to the equilibrium probabilities for the true transition states (i.e., conformations with the proper side chain mode). The correction not only increased the barrier height, but also modified the barrier location, with the peak at $\alpha^* = 0.61$ for the corrected free energy.

The local diffusion coefficients at each umbrella window, as shown in Fig. 3B, were calculated from umbrella sampling simulations (of 5 ns each) without Hamiltonian replica exchange, as described in Methods. These data points were averaged into a smooth curve representing the position-dependent diffusion coefficients $D(\alpha)$. The $D(\alpha)$ curve (Fig. 3B) does not exhibit significant variations across the range of α in this case, implying that a constant diffusion coefficient would probably be sufficient to describe the transition.

Because the dynamics at the barrier top α^* is a major determinant for the kinetic rates of the transition, we released the system at α^* in 200 independent unbiased simulations and observed their free evolution over a simulation time of 1 ns, as described in Methods. The initial coordinates of these simulations were chosen such that the side chains were not in the false transition states discussed earlier. Among the 200 simulations starting at α^*, the reaction coordinate in 111 of them ended up with values smaller than α^* after 1 ns, whereas 89 simulations finished with the reaction coordinate larger than α^*. The splitting probabilities

(55.5% vs. 44.5%) here are in good agreement with the ideal values (50% vs. 50%), thus validating that our α^* as the peak of $G(\alpha)$ is indeed close to the real barrier for the transition. Furthermore, as described in Methods, the 200 unbiased simulations form 100 pairs, with the two simulations in each pair having identical initial coordinates but opposite initial velocities for all the atoms. If the two simulations commit to the two metastable states on opposite sides of the barrier, they can be pieced into a continuous "reactive" trajectory that represents a spontaneous transition between the two states. If the barrier crossing is highly diffusive and each conformation at α^* has equal probabilities of committing to the two sides of the barrier, each pair of simulations should have a probability of 0.5 to commit to opposite states (thus forming a reactive trajectory). Among the 100 pairs of simulations here, 21 of them committed to either sides of α^* at the end of 1 ns whereas 79 committed to the same side. Although the ratio (0.21) is smaller than the ideal value (0.5), it nonetheless indicates that a substantial portion of the conformations at α^* is indeed on the transition pathway between OF and OF'.

In principle, from the free evolutions starting at the barrier top, one could compute the transmission coefficient[48] and subsequently the transition rates.[48] However, some of the 200 simulations here still stayed near the barrier region without fully committing at the end of 1 ns. We also observed re-crossing of the barrier in several simulations. These observations suggest that the barrier crossing here is diffusive and the transmission coefficient would be very small. Consequently, a direct calculation of the transmission coefficient would bear very large statistical errors. Instead, it appears to be more feasible to characterize the transition kinetics by a diffusion model. As discussed earlier, the calculated diffusion coefficient along the reaction coordinate α does not exhibit large variations across the entire sampled range. Therefore one could reasonably assume a constant diffusion coefficient *D* for the transition, and the 200

unbiased simulations can be used for a better estimate of this effective diffusion coefficient.[36] Specifically, given any value of D, we can numerically solve the Smoluchowski equation (Eq. 22) to obtain the probability distribution of α after a free evolution of 1 ns from α^*. We then identify the D value that gives rise to the optimal match between the predicted distribution and the actual histogram of α over the last frames of the 200 simulations.[36] In our case the best-fit value is 13 /μs for the diffusion coefficient, as shown in Fig. 3B (*dashed* horizontal line), which is smaller by roughly a factor of 4 than the average diffusion coefficient calculated from the simulations with harmonic restraints. The constant diffusion coefficient obtained from the unbiased simulations here was adopted in the rate calculations below, as it better describes the observed dynamics at the barrier top.

With the free energy $G(\alpha)$ and the effective diffusion coefficient D obtained, we then calculated the thermodynamics and kinetics for the transition. Using Eq. 17, the equilibrium probabilities for the OF and the OF' states are ~1 and ~2×10^{-5}, respectively, indicating that the OF' state has an integrated free energy 6.5 kcal/mol higher than the OF state. For the spontaneous transition rate k_0 in the equilibrium ensemble, a numerical integration of Eq. 23 yields $1/k_0 \sim$ 9 ms. According to Eq. 24, the forward and backward transitions rates are thus given by $1/k_{OF\to OF'} \sim$ 9 ms and $1/k_{OF'\to OF} \sim$ 0.2 μs.

DISCUSSION

In this study, we demonstrated a practical approach, based on the finite temperature string method, to obtain a pathway and a 1D free energy for transitions between two metastable

conformations. The transition pathway is a curve in the configuration space formed by coarse coordinates, and the free energy is a function of the curve parameter that serves as the reaction coordinate for the transition. Our approach is a natural generalization of the well-established 1D umbrella sampling technique to the problem of sampling along a curved pathway.[18] The method is conceptually simple and practically convenient to implement. More importantly, many useful tools for the traditional umbrella sampling, such as Hamiltonian replica exchange[20] and error analysis[30], become directly applicable here to improve the efficiency or to detect potential sampling problems such as hysteresis[30]. Furthermore, once the protein dynamics can be mapped to the time evolution of the reaction coordinate, we may employ a diffusion model to calculate the forward and backward transition rates, which are the major kinetic quantities to describe the transition. The approach here should be generally applicable for elucidating transitions between known protein conformations.

We applied our approach to characterize a side chain transition in the Mhp1 transporter, and obtained a transition pathway and a 1D free energy from the umbrella sampling simulations with Hamiltonian replica exchange[20]. The free energy describes the equilibrium probability distribution (see Eq. 5) when the entire protein conformation is mapped to the single reaction coordinate. In the meantime, some highly distinct conformations, such as the two side chain modes discussed earlier, may be on the same hyperplane and mapped to an identical value of the reaction coordinate. In our case here, some conformations apparently at the free energy barrier are actually false transition states and are not on the pathway of real transitions. These observations indicate that our chosen reaction coordinate is not a perfect one. In retrospect, incorporating the orientation of the Met178:C_γ-Phe305:C_4 vector into the coarse coordinates would help define a reaction coordinate that better distinguishes the true transition states. In this

study, alternatively, we chose to identify and exclude the false transition states through a posterior analysis of the simulation trajectories. Exclusion of those false transition states resulted in an improved free energy which better describes the equilibrium probabilities of real transition intermediates in the barrier region. Free dynamics at the identified barrier top after the correction verified that the conformations there are reasonable transition states, as a number of reactive trajectories can indeed be produced by releasing the system in a pair of unbiased simulations with opposite atomic velocities. Finally, with a diffusion coefficient estimated from the unbiased simulations, the kinetic rates for the transition were obtained from the diffusion model. We note that the collective simulation time in this study is about 10 μs, much shorter than the millisecond sampling times required to observe spontaneous transitions in this system, thus demonstrating that our approach can be much more efficient than straightforward simulations.

The finite temperature string method was originally implemented in simulations with the reaction coordinate constrained at constant values.[14] Recently, simulations in non-overlapping Voronoi cells[15, 16] have become a more common implementation of the method. When the principal curve[15] in this method is sufficiently straight, the confinement in a Voronoi cell is similar to applying a flat 1D potential that rises to infinity at the two boundaries. If the reaction coordinate is sufficiently good to ensure fast equilibration of all orthogonal motions such that the time evolution of the reaction coordinate can be considered a Markov process, the simulations in the Voronoi cells could directly provide not only the thermodynamics of the conformational space, but also the transition rates[49]. In contrast, our 1D umbrella sampling with harmonic restraining potentials, as similarly adopted previously,[18] is primarily designed to calculate the equilibrium probabilities, and need to be supplemented with other information such as the diffusion coefficients or the transmission coefficient to further obtain the kinetic rates. In this

aspect, our method appears to be less advantageous compared to the Voronoi implementation if a perfect reaction coordinate is adopted for the sampling.

When the reaction coordinate is not perfect, however, our approach can be practically more reliable than the Voronoi implementation. A particular advantage of using overlapping potentials in the umbrella sampling comes from the Hamiltonian replica exchange[20], which has been demonstrated to accelerate convergence[50]. In our case here, e.g., a single value of the reaction coordinate could correspond to two distinct conformational modes, and spontaneous interconversions between the two modes would be practically impossible if the simulation is confined in a Voronoi cell. Consequently, the simulation will only sample one mode as determined by the initial coordinates, without ever visiting the other mode, and the sampling result will thus bear hysteresis as it depends on the initial condition. In contrast, Hamiltonian replica exchange allowed our simulations to properly sample both modes in the same umbrella window. With the correct thermodynamics obtained, we could then exclude the irrelevant conformational mode and obtain the probabilities of the true transition states. Furthermore, the exchange rates of the replicas could indicate slow equilibration in degrees of freedom orthogonal to the reaction coordinate.[50] In addition, using overlapping umbrella potentials allows one to check the consistency[30] (Fig. 3C) between the results from neighboring umbrella windows, which could also reveal potential sampling problems[30].

For complex protein conformational changes, it is challenging to identify all relevant degrees of freedom[29] and normally very difficult to find a perfect reaction coordinate a priori. Therefore, it is typically not safe to assume fast equilibration for all orthogonal coordinates in the high-dimensional conformational space. In such cases, methods that are more tolerant to somewhat problematic reaction coordinates should be preferable in practice, and it is also

particularly important to examine potential sampling problems such as hysteresis. As demonstrated here, our method of umbrella sampling with Hamiltonian replica exchange could reliably calculate the free energy even when the reaction coordinate is not perfect. The method also comes with convenient tools to estimate the statistical errors and examine the consistency and convergence of the calculation. Therefore, our approach here is a practical choice for characterizing complex conformational transitions.

The side chain flipping of Phe305 in the transition from the OF to the OF' state here is merely the first step in the complete OF-to-IF transition. Going from the OF' further to the IF state, a number of other major changes need to occur, both in the overall protein conformation and in the local secondary structures[29]. Without the side chain flipping here, however, the OF-to-IF transition would be energetically unfavorable because the Phe305 side chain could sterically hinder the closing of the extracellular pocket when the IF conformation is approached. Our results indicate that the OF state is thermodynamically more stable than the OF' state, by a free energy difference of ~6.5 kcal/mol. At equilibrium, the population of the OF' state would thus be lower than the OF population by almost five orders of magnitude. One therefore would not expect to capture the Phe305 side chain in the flipped-out position in the OF crystal structures[22, 24] of Mhp1. Kinetically, our calculations predict that the spontaneous transition from OF to OF' occurs at the millisecond time scale. It is thus not surprising that no spontaneous flipping of this side chain was reported in previous MD simulations. Reversely, transitions from the OF' to the OF state is much faster, at the sub-microsecond time scale. This implies that once the OF' state is reached during the IF-to-OF transition, the Phe305 side chain will flip to the position in the OF state in a relatively short time.

The equilibrium distance distributions for some pairs of spin-labeled residues in Mhp1 were measured by double electron-electron resonance spectroscopy.[28] These distance measurements generally implied the presence of intermediate metastable conformations in addition to the OF and the IF states. Our simulations suggest OF' as an intermediate state, but the thermodynamics and kinetics for the transitions between the OF' and the IF states are not known yet. Therefore, it is currently unclear whether the side chain flipping studied here would be a rate-limiting step in the complete transition between OF and IF, and a direct comparison to experiments is not ready at this stage.

CONCLUSION

This study concerns rare transitions between two metastable protein conformations. To computationally characterize such transitions, it is vital to incorporate all relevant degrees of freedom into the coarse coordinates representing the conformation. Then the transition pathway can be described by a curve in the configuration space formed by the coarse coordinates. Our umbrella sampling approach presented here could be applied to refine a transition pathway and to calculate a 1D free energy as a function of the curve parameter that serves as the reaction coordinate. The obtained free energy is directly related to the marginal probability distribution that integrates out all degrees of freedom orthogonal to the reaction coordinate, and can predict the equilibrium probabilities of the two metastable conformations. If the transition can be considered a 1D diffusion process along the reaction coordinate, the kinetic rates may also be estimated from the diffusion coefficient and the free energy.

For the Mhp1 transporter to undergo a complete transition from the OF to the IF conformation, our previous simulations[29] suggested that the first step would be the flipping of the Phe305 side chain. By employing umbrella sampling, our simulations here further indicate that the flipping will result in an intermediate conformation OF' with substantially higher free energy than the OF state. By applying the computational methods demonstrated here, it should be possible to similarly elucidate other transition steps toward the IF state, which will eventually allow the prediction for the overall transition rates between the OF and the IF conformations as well as the effects of potential mutations.

ACKNOWLEDGEMENT

This work was supported by Grants to Enhance Interdisciplinary Research and Education (GEIRE) from Institute for Mathematical Modeling and Computational Science, School of Science, IUPUI. The simulations were performed on the Big Red II supercomputer at Indiana University as well as a Linux cluster at School of Science, IUPUI.

References

1. Khalili-Araghi, F.; Gumbart, J.; Wen, P. C.; Sotomayor, M.; Tajkhorshid, E.; Schulten, K. Molecular Dynamics Simulations of Membrane Channels and Transporters. *Curr. Opin. Struct. Biol.* **2009,** *19*, 128-137.
2. Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010,** *330*, 341-346.
3. Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011,** *334*, 517-520.
4. Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002,** *53*, 291-318.
5. Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. Targeting Biomolecular Flexibility with Metadynamics. *Curr. Opin. Struct. Biol.* **2010,** *20*, 148-154.
6. Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004,** *120*, 11919-11929.
7. Darve, E.; Rodriguez-Gomez, D.; Pohorille, A. Adaptive Biasing Force Method for Scalar and Vector Free Energy Calculations. *J. Chem. Phys.* **2008,** *128*, 144120.
8. Faradjian, A. K.; Elber, R. Computing Time Scales from Reaction Coordinates by Milestoning. *J. Chem. Phys.* **2004,** *120*, 10880-10889.
9. Perilla, J. R.; Beckstein, O.; Denning, E. J.; Woolf, T. B. Computing Ensembles of Transitions from Stable States: Dynamic Importance Sampling. *J. Comput. Chem.* **2011,** *32*, 196-209.
10. Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophys. J.* **1996,** *70*, 97-110.

11. Isralewitz, B.; Gao, M.; Schulten, K. Steered Molecular Dynamics and Mechanical Functions of Proteins. *Curr. Opin. Struct. Biol.* **2001,** *11*, 224-230.

12. E, W.; Ren, W.; Vanden-Eijnden, E. String Method for the Study of Rare Events. *Phys. Rev. B* **2002,** *66*, 52301.

13. Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String Method in Collective Variables: Minimum Free Energy Paths and Isocommittor Surfaces. *J. Chem. Phys.* **2006,** *125*, 24106.

14. E, W.; Ren, W.; Vanden-Eijnden, E. Finite Temperature String Method for the Study of Rare Events. *J. Phys. Chem. B* **2005,** *109*, 6688-6693.

15. Vanden-Eijnden, E.; Venturoli, M. Revisiting the Finite Temperature String Method for the Calculation of Reaction Tubes and Free Energies. *J. Chem. Phys.* **2009,** *130*, 194103.

16. Ovchinnikov, V.; Karplus, M.; Vanden-Eijnden, E. Free Energy of Conformational Transition Paths in Biomolecules: The String Method and Its Application to Myosin Vi. *J. Chem. Phys.* **2011,** *134*, 085103.

17. Kirmizialtin, S.; Johnson, K. A.; Elber, R. Enzyme Selectivity of Hiv Reverse Transcriptase: Conformations, Ligands, and Free Energy Partition. *J. Phys. Chem. B* **2015,** *119*, 11513-11526.

18. Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. A Recipe for the Computation of the Free Energy Barrier and the Lowest Free Energy Path of Concerted Reactions. *J. Phys. Chem. B* **2005,** *109*, 6676-6687.

19. Song, H.; Zhu, F. Conformational Dynamics of a Ligand-Free Adenylate Kinase. *PLoS One* **2013,** *8*, e68023.

20. Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002,** *116*, 9058-9067.

21. Suzuki, S.; Henderson, P. J. The Hydantoin Transport Protein from Microbacterium Liquefaciens. *J. Bacteriol.* **2006,** *188*, 3329-3336.

22. Weyand, S.; Shimamura, T.; Yajima, S.; Suzuki, S.; Mirza, O.; Krusong, K.; Carpenter, E. P.; Rutherford, N. G.; Hadden, J. M.; O'Reilly, J.; et al. Structure and Molecular Mechanism of a Nucleobase-Cation-Symport-1 Family Transporter. *Science* **2008,** *322*, 709-713.

23. Shimamura, T.; Weyand, S.; Beckstein, O.; Rutherford, N. G.; Hadden, J. M.; Sharples, D.; Sansom, M. S.; Iwata, S.; Henderson, P. J.; Cameron, A. D. Molecular Basis of Alternating Access Membrane Transport by the Sodium-Hydantoin Transporter Mhp1. *Science* **2010,** *328*, 470-473.

24. Simmons, K. J.; Jackson, S. M.; Brueckner, F.; Patching, S. G.; Beckstein, O.; Ivanova, E.; Geng, T.; Weyand, S.; Drew, D.; Lanigan, J.; et al. Molecular Mechanism of Ligand Recognition by Membrane Transport Protein, Mhp1. *EMBO J.* **2014,** *33*, 1831-1844.

25. Adelman, J. L.; Dale, A. L.; Zwier, M. C.; Bhatt, D.; Chong, L. T.; Zuckerman, D. M.; Grabe, M. Simulations of the Alternating Access Mechanism of the Sodium Symporter Mhp1. *Biophys. J.* **2011,** *101*, 2399-2407.

26. Li, J.; Shaikh, S. A.; Enkavi, G.; Wen, P. C.; Huang, Z.; Tajkhorshid, E. Transient Formation of Water-Conducting States in Membrane Transporters. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110*, 7696-7701.

27. Zhao, C.; Noskov, S. Y. The Molecular Mechanism of Ion-Dependent Gating in Secondary Transporters. *PLoS Comput. Biol.* **2013,** *9*, e1003296.

28. Kazmier, K.; Sharma, S.; Islam, S. M.; Roux, B.; McHaourab, H. S. Conformational Cycle and Ion-Coupling Mechanism of the Na+/Hydantoin Transporter Mhp1. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111*, 14752-14757.

29. Song, H.; Zhu, F. Conformational Changes in Two Inter-Helical Loops of Mhp1 Membrane Transporter. *PLoS One* **2015,** *10*, e0133388.

30. Zhu, F.; Hummer, G. Convergence and Error Estimation in Free Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **2012,** *33*, 453-465.

31. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J. Comput. Chem.* **1992,** *13*, 1011-1021.

32. Khavrutskii, I. V.; Arora, K.; Brooks, C. L., 3rd Harmonic Fourier Beads Method for Studying Rare Events on Rugged Energy Surfaces. *J. Chem. Phys.* **2006,** *125*, 174108.

33. Zhu, F.; Hummer, G. Pore Opening and Closing of a Pentameric Ligand-Gated Ion Channel. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107*, 19814-19819.

34. Peters, B. Reaction Coordinates and Mechanistic Hypothesis Tests. *Annu. Rev. Phys. Chem.* **2016,** *67*, 669-690.

35. Hummer, G. Position-Dependent Diffusion Coefficients and Free Energies from Bayesian Analysis of Equilibrium and Replica Molecular Dynamics Simulations. *New J. Phys.* **2005,** *7*, 34.

36. Zhu, F.; Hummer, G. Theory and Simulation of Ion Conduction in the Pentameric Glic Channel. *J. Chem. Theory Comput.* **2012,** *8*, 3759-3768.

37. Flyvbjerg, H.; Petersen, H. G. Error-Estimates on Averages of Correlated Data. *J. Chem. Phys.* **1989,** *91*, 461-466.

38. MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998,** *102*, 3586-3616.

39. MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* **2004,** *126*, 698-699.

40. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr. Optimization of the Additive Charmm All-Atom Protein Force Field Targeting Improved Sampling of the Backbone Phi, Psi and Side-Chain Chi(1) and Chi(2) Dihedral Angles. *J. Chem. Theory Comput.* **2012,** *8*, 3257-3273.

41. Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D., Jr.; Pastor, R. W. Update of the Charmm All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010,** *114*, 7830-7843.

42. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.

43. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with Namd. *J. Comput. Chem.* **2005,** *26*, 1781-1802.

44. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes *J. Comput. Phys.* **1977,** *23*, 327-341.

45. Miyamoto, S.; Kollman, P. A. Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992,** *13*, 952-962.

46. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.

47. Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J. Chem. Phys.* **1995,** *103*, 4613-4621.

48. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications;* Academic Press: San Diego, 2002.

49. Vanden-Eijnden, E.; Venturoli, M. Markovian Milestoning with Voronoi Tessellations. *J. Chem. Phys.* **2009,** *130*, 194101.

50. Neale, C.; Madill, C.; Rauscher, S.; Pomes, R. Accelerating Convergence in Molecular Dynamics Simulations of Solutes in Lipid Membranes by Conducting a Random Walk Along the Bilayer Normal. *J. Chem. Theory Comput.* **2013,** *9*, 3686-3703.

51. Humphrey, W.; Dalke, A.; Schulten, K. Vmd: Visual Molecular Dynamics. *J. Mol. Graph.* **1996,** *14*, 33-38.

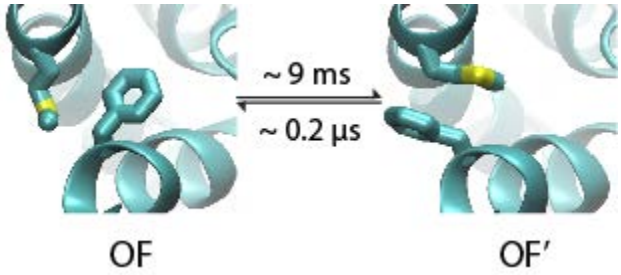
FIGURE LEGENDS

Figure 1. The OF and IF crystal structures[22, 23] and the OF' conformation of Mhp1 viewed from the extracellular side. The OF' state is highly similar to OF, except that the side chain of Phe305 is flipped to the outside of the protein as in the IF state. The extracellular OUT7-8 helix and the EL4 loop are made transparent to display the side chains of Met178 and Phe305 underneath. The molecular images were rendered using VMD.[51]

Figure 2. The positions of the C_4 atom of Phe305 relative to the C_γ atom of Met178 in the umbrella sampling simulations. With the coordinates of the two atoms denoted as $\vec{r}_2$ and $\vec{r}_1$, respectively, every dot in the figure represents the position $\Delta\vec{r} \equiv \vec{r}_2 - \vec{r}_1$ from one frame of the simulation trajectories, and the sphere at the center represents the origin, or the position of the Met178:C_γ atom. The continuous curve is the trace of $\Delta\vec{r}$ along the pathway curve for the umbrella sampling. The molecular images show the side chains of Met178 and Phe305 as well as part of the TM5 and TM8 backbones for four conformations. The Phe305:C_4 and the Met178:C_γ atoms, which were included in the coarse coordinates in this study, are shown as small spheres, with other side chain atoms shown in the licorice representation.[51] The displayed conformations include two snapshots in the barrier region of the transition as well as the OF and the OF' structures. The two snapshots correspond to similar values of the reaction coordinate, but have very different positions for the side chains. The conformation in one of the snapshots cannot directly evolve to the OF structure and is a false transition state.

Figure 3. Free energy and diffusion coefficients as a function of the reaction coordinate α. (**A**) The original free energy profile $G_0(\alpha)$ calculated from the umbrella sampling trajectories is shown by the *dashed* line, and the profile $G(\alpha)$ after correcting for the false transition states in the barrier region (as explained in the text) is shown by the *solid* line. (**B**) Each *star* represents

the calculated local diffusion coefficient at an umbrella window. The *solid* curve is obtained by smoothing these data points by a moving average over 11 windows. The *dashed* horizontal line shows the constant effective diffusion coefficient estimated from the 200 unbiased simulations by a maximum-likelihood approach described in the text. (**C**) The inconsistency coefficients[30] for pairs of histograms in adjacent umbrella windows.



TOC graphic